ChemPhysChem

Accepted Article

Interpretable Machine Learning Models for Practical Antimonate Electrocatalyst Performance

Shyam Deo,[†‡] Melissa Kreider,[†‡] Gaurav Kamat,[†‡] McKenzie Hubert,[†‡]

José A. Zamora Zeledón,[†‡] Lingze Wei,[†‡] Jesse Matthews,[†‡] Nathaniel Keyes,[†‡]

Ishaan Singh,[†‡] Thomas Jaramillo,[†‡] Frank Abild-Pedersen,[‡]

Michaela Burke Stevens,[‡∥] Kirsten Winther,[‡§] Johannes Voss[‡¶*]

[†]*Department of Chemical Engineering, Stanford University, Stanford, CA 94305, United States*
[‡]*SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, United States*

[∥]mburkes@slac.stanford.edu
[§]winther@slac.stanford.edu
[¶]vossj@slac.stanford.edu
[*]Corresponding Author

This article is dedicated to Professor Jens Nørskov.

Abstract

Computationally predicting the performance of catalysts under reaction conditions is a challenging task due to the complexity of catalytic surfaces and their evolution in situ, different reaction paths, and the presence of solid-liquid interfaces in the case of electrochemistry. We demonstrate here how relatively simple machine learning models can be found that enable prediction of experimentally observed onset potentials. Inputs to our model are comprised of data from the oxygen reduction reaction on non-precious transition-metal antimony oxide nanoparticulate catalysts with a combination of experimental conditions and computationally affordable bulk atomic and electronic structural descriptors from density functional theory simulations. From human-interpretable genetic programming models, we identify key experimental descriptors and key supplemental bulk electronic and atomic structural descriptors that govern trends in onset potentials for these oxides and deduce how these descriptors should be tuned to increase onset potentials. We finally validate these machine learning predictions by experimentally confirming that scandium as a dopant in nickel antimony oxide leads to a desired onset potential increase. Macroscopic experimental factors are found to be crucially important descriptors to be considered for models of catalytic performance, highlighting the important role machine learning can play here even in the presence of small datasets.

1

1. Introduction

Active, selective, and stable catalysts for the oxygen reduction reaction (ORR) are essential for efficient fuel cell technology. Platinum-based systems are currently the most efficient ORR catalysts, and the discovery of catalyst materials consisting of non-precious metals and other earth-abundant and inexpensive elements is an active field of research.[1–3] Here, we focus on first-row transition-metal antimony oxides as one such class of catalyst materials with promising ORR performance.[4,5] We aim to develop machine learning (ML) models that can be used to optimize catalyst performance by combining experiments with atomic-level computer simulations for model training and validation, using the ORR on transition-metal antimony oxides as a prototype catalytic reaction in this study.

Trends in heterogeneous catalytic activity are readily rationalized by electronic descriptor models, such as the d-band model by Nørskov and coworkers for transition-metal surface reactivity,[6] and p-band,[7,8] crystal orbital Hamiltonian population,[9] and oxidation state models[10] for transition-metal oxide reactivity. These electronic descriptors are typically extracted from bulk or idealized surface density functional theory (DFT) simulations. The complexity of real oxide electrocatalysts in terms of the electrochemical interface, surface structure, oxygen vacancies, and support effects render a purely atomistic-based first-principles prediction of catalyst performance (i.e., catalytic activity, selectivity, and stability) challenging. Empirical models[11–14] aim to predict catalyst performance under such experimental or realistic conditions as well as to potentially allow for an optimization of e.g. synthesis conditions or catalyst composition. Stochastic techniques such as design of experiments and Bayesian optimization machine-learning approaches can be applied for optimization of such catalyst parameters.[15–17] Interpretable machine-learning approaches furthermore allow for human rationalization of the discovered models for catalyst performance and could lead to understanding of reaction mechanism under realistic conditions.[18,19]

Due to the limited quantity of experimental data relating catalyst performance to catalyst composition, structure, and reaction conditions, we focus here on two complementary machine-learning approaches that can deal with sparse training and validation data. Firstly, we apply the genetic

programming[20] technique of symbolic regression,[21] where mathematical equations are evolved as programs with mutations and crossover of terms between programs. Symbolic regression (by means of genetic programming and other ML approaches aiming at discovering symbolic expressions[22–24]) has been used successfully to discover (or re-discover) mathematical models from data.[25–28] In heterogeneous catalysis, symbolic regression has been applied to identify descriptors for binding energies and reactivities.[29,30] The resulting ML models in (relatively) simple mathematical form have the advantage of allowing for human interpretation and rationalization. Secondly, we apply the technique of Gaussian process regression, [31] which in the context of computational chemistry can be used to regress on atomistic properties.[32]

We show how both approaches yield models for ORR onset potentials with root mean-square errors (RMSE) of 30-50 mV. From the simple mathematical form of the genetic-programming models for ORR onsets, we identify atomic and electronic structure descriptors that can be tuned for optimizing performance. We demonstrate the predictive power of these descriptors with experiments on scandium-doped nickel antimony oxide, yielding a modest increase in ORR onset potential.
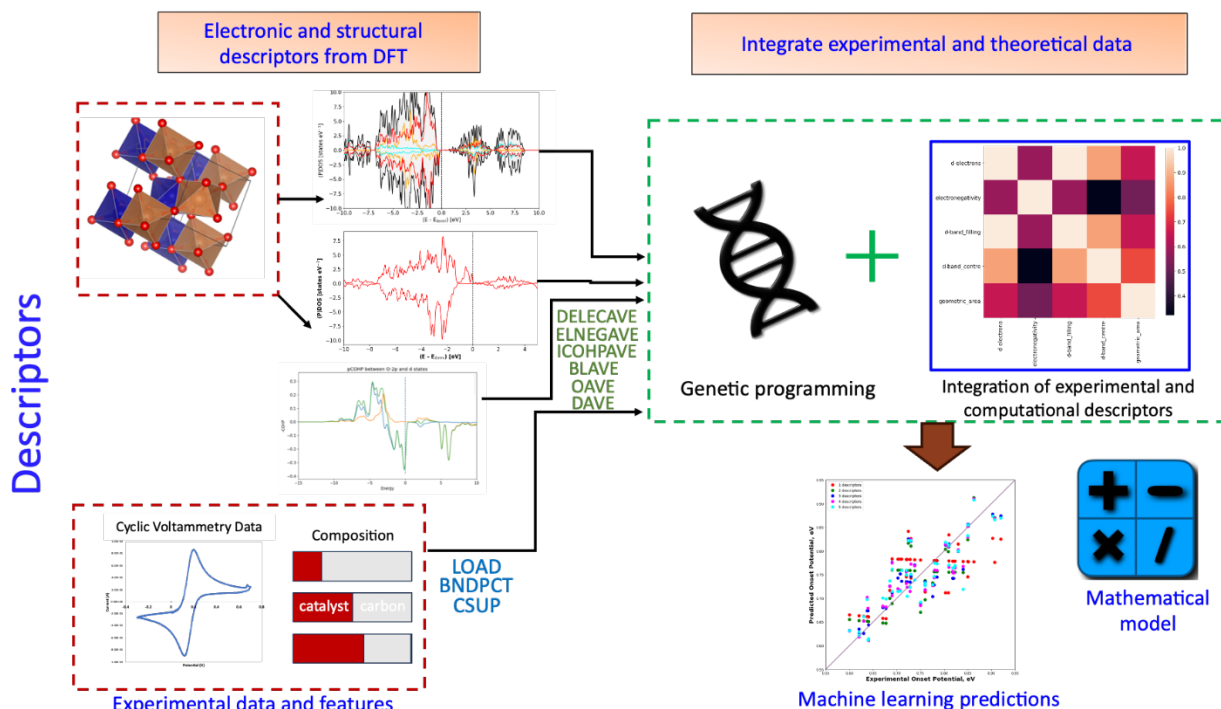
Figure 1. Schematic for the experimental and theoretical data integration approach. The schematic comprises four primary elements: gathering experimental data on ORR catalyst performance, which includes data from prior studies and additional experiments conducted; computation of electronic descriptors for bulk transition-metal antimony oxides using DFT to complement and enhance the experimental data; storing and labeling both experimental and computational data in our database (catalysis-hub.org); and conducting a search for ML models.

2. Methods

Our approach (see Figure 1 for an overview) consists of four major components: the collection of experimental data of ORR catalyst performance (stemming here from previously published work[4] as well as additional experiments performed for this project), the computation of bulk transition-metal antimony oxide electron descriptors using DFT to be integrated with and supplement the experimental data, the integrated storage and labeling of experimental and computational data on our database catalysis-hub.org,[33] and the ML model search.

2.1. Experimental data collection

In this study, we confine the materials and model search to the stoichiometry of the promising antimonate Mn(SbO$_3$)$_2$. This allows us to focus this study on models that describe changes to the performance of this ORR catalyst by means of alloying and doping at the transition-metal sites without having to introduce more complex descriptors that capture the stoichiometry of the system. While we consider Cr as a dopant, we exclude chromium antimonate (i.e. the extreme case of 100% Cr-doping) from our search, as this material displays a different stoichiometry (CrSbO$_4$).[34] The majority of experimental data used to create the ML models was previously collected, characterized, and published for the Mn, Fe, Ni, and Co antimonates.[4] To test the predictions herein, new catalysts containing Sc as well as select samples for extended loading studies were synthesized, characterized, and reported exclusively for and in this article as described below.

2.1.1. Antimony oxide synthesis

Transition-metal antimony oxides (TM antimonates) are synthesized[4] as nanoparticles via a colloidal scheme[35] in reaction grade anhydrous ethanol. Transition-metal nitrate salts (such as Mn(NO$_3$)$_2$ and Sc(NO$_3$)$_3$, Sigma Aldrich) are dispersed in ethanol in molar ratios according to the doping amounts. Separately, a solution of antimony chloride (SbCl$_3$, Sigma Aldrich) is prepared with ethanol and a third solution of ethylenediamine in ethanol is prepared. The three solutions are combined dropwise, first by adding the antimony chloride to the metal nitrates, and second by adding the ethylenediamine to the metal nitrate-antimony mixture. The solution is left stirring at room temperature for 24 hours and then transferred to a ceramic boat for drying in a furnace under air at 200C. The dry powder is then ground via mortar and pestle and transferred to another furnace boat for heat treatment at 800C for 6 hours in air. The resulting oxide powder is again ground via mortar and pestle and set aside for ink making.

2.1.2. Electrode preparation and electrochemical testing

The inks are prepared using a fixed mass of the oxide nanoparticle powder, a solution of 10 wt% Nafion^{TM} in higher alcohols, Vulcan XC-72 carbon, and isopropanol. The amount and ratio of the constituents are

varied to achieve a specific catalyst loading and can be calculated from the descriptors used in the data model such as conductive support to catalyst ratio, mass loading of catalyst, and binder percentage with the definitions in Table 1. These materials are combined in a vial and sonicated for 15 minutes prior to deposition to ensure sufficient dispersion. The inks are deposited on polished glassy carbon disk electrodes for electrochemical testing on rotating disk electrode (RDE) and selectivity measurement on rotating ring-disk electrode (RRDE). Samples are measured at 1600rpm in oxygen- or nitrogen-saturated potassium hydroxide or perchloric acid electrolyte. In the case of RRDE, an Au ring is used in alkaline electrolyte and a Pt disk is used in acidic electrolyte with both held at 1.2 V vs RHE throughout the experiment.

2.1.3. Material characterization

Physical characterization is done with x-ray diffraction (XRD, Malvern Panalytical Empyrean, iCore Optics, GaliPIX3D Detector) and x-ray photoelectron spectroscopy (XPS, PHI Instruments VersaProbe 3 with Ar^+ and electron beam neutralization). All material characterization is performed before testing of the nanoparticle oxide powders. For simplicity of the ML models, the as-synthesized catalyst is considered the active catalyst in situ. While some level of material dynamics is expected in any catalytic system, based on in situ and post characterization measurements done previously on a subset of this material set, minimal dissolution and bulk structural or surface oxidation state changes are expected for this system.[5] The nominal structural integrity throughout testing makes this system ideal for ML modeling as it allows for the reasonable exclusion of material stability in the model.

2.2. Computational Methods

Owing to the complexity of electrochemical reactions on oxides, we focus the simulation work on computing bulk descriptors using DFT. The purpose of these bulk properties is to serve as additional descriptors for the experimentally studied catalysts. The combination of bulk DFT and experimental descriptors is used as input to ML models trained to reproduce experimentally observed ORR onset potentials.

6

2.2.1. Electronic structure methods

Spin polarized plane-wave density functional theory (DFT) calculations are performed using the Vienna ab initio Simulation Package (VASP) version 5.4.4,[36] using the atomic simulation environment (ASE) to setup and control simulations.[37] Electronic exchange and correlation energies are approximated using the PBE functional.[38] The projector-augmented wave (PAW) method[39] is employed to describe the ionic cores, and a kinetic energy cutoff of 520 eV is used for the plane wave basis sets. The Manganese (Mn), Nickel (Ni), Iron (Fe), Chromium (Cr) and Scandium (Sc) 3p, 3d and 4s electrons, the Cobalt (Co) 3d and 4s, the Antimony (Sb) 5s and 5p, and the Oxygen (O) 2s and 2p electrons are treated as valence electrons for the calculations, using PAW pseudopotentials provided with the VASP code.[40] Geometries are optimized with a force convergence criterion of 0.02 eV/Å, and total energies are converged to at least 10^{-5} eV. Electronic occupation broadening with a Gaussian with a width of 0.05 eV is used for all calculations. Brillouin zone sampling is performed with grid spacings of about 0.3 Å^{-1}. To correct for self-interaction errors at the transition-metal ion sites, Hubbard-U corrections are applied within the DFT+U approach.[41,42] We adopt the Hubbard parameter strengths from the Materials Project[43] with U-values for the 3d-states of Cr, Mn, Fe, Co, and Ni of 3.5, 3.9, 5.3, 3.32, and 6.2 eV, respectively. Only structural relaxations are performed with Hubbard U-corrections, while subsequent single-point calculations at fixed crystal structure for electronic descriptor extraction are performed without these Hubbard corrections. We have found that including the Hubbard-U failed to open the electronic band gap for several of the antimony oxides considered. Instead, for these oxides, metallic density of states remains at the Fermi level, and the Hubbard terms spuriously push valence states down in energy (by an energy shift on the order of U) leading to significant offsets in electronic descriptors such as band centers for these (spuriously) metallic systems (see Figure S1c). We found that breaking the antiferromagnetic symmetry in the antimony oxides (either by residual magnetic moments present on oxygen or by metal-site doping) exacerbated the unphysical shift in the band structures when introducing a Hubbard-U, and we refrained from a large super cell search with dopant-site magnetic order compatible with band gaps.

In the case of metallic DOS, Hubbard corrections should not be applied due to the induced spurious valence band shifts. We therefore choose to compute all band structures without Hubbard corrections, so that we have a consistent set of electronic descriptors with well-defined relative shifts for pure, alloyed, and doped transition-metal antimony oxides. We refer to the Supporting Information and Figure S1 for an in-depth discussion.

DOS calculations for the bulk structures are performed at fixed charge density but with finer k-point mesh (0.2 $Å^{-1}$ spacing) and using the tetrahedron method[44] following a self-consistent single-point DFT calculation for the same system. From these non-self-consistent simulations, atom-site projected DOS and the resulting band centers are computed.

Integrated Crystal Orbital Hamilton Populations (ICOHP) are calculated using the LOBSTER (Local-Orbital Basis Suite Towards Electronic-Structure Reconstruction) code version 4.0.0.[45] The integration of the COHP is performed over metal ($3d$) - oxygen ($2p$) orbitals up to the Fermi level, including all transition-metal-oxygen pairs in the unit cell.

2.2.2. Bulk Model Structures

Transition-metal (M) antimonates: All transition-metal (M) antimonates in our experimental dataset of materials, $MnSb_2O_6$, $NiSb_2O_6$, $CoSb_2O_6$, and $FeSb_2O_6$ are optimized with bulk structures in the tetragonal $P4_2/mnm$ space group for all oxides, using the structure of $MnSb_2O_6$ (mp-763546) from the Materials Project Database (https://materialsproject.org/)[43] as a prototype. We use the P42/mnm structure as the prototype due to its small primitive cell size, neglecting small changes in local coordination in other phases. An optimized primitive cell of the bulk structure ($1 \times 1 \times 1$ super cell size) of the antimonates is considered for further electronic and structural analysis (Figure 1a-b). The primitive unit cell consists of 2 metal (M) sites, 4 antimony (Sb) and 12 oxygen atoms. In accordance with the reported experimental structures in the database, an antiferromagnetic (AFM) structure is considered, and the lowest energy structure is taken for further analysis.
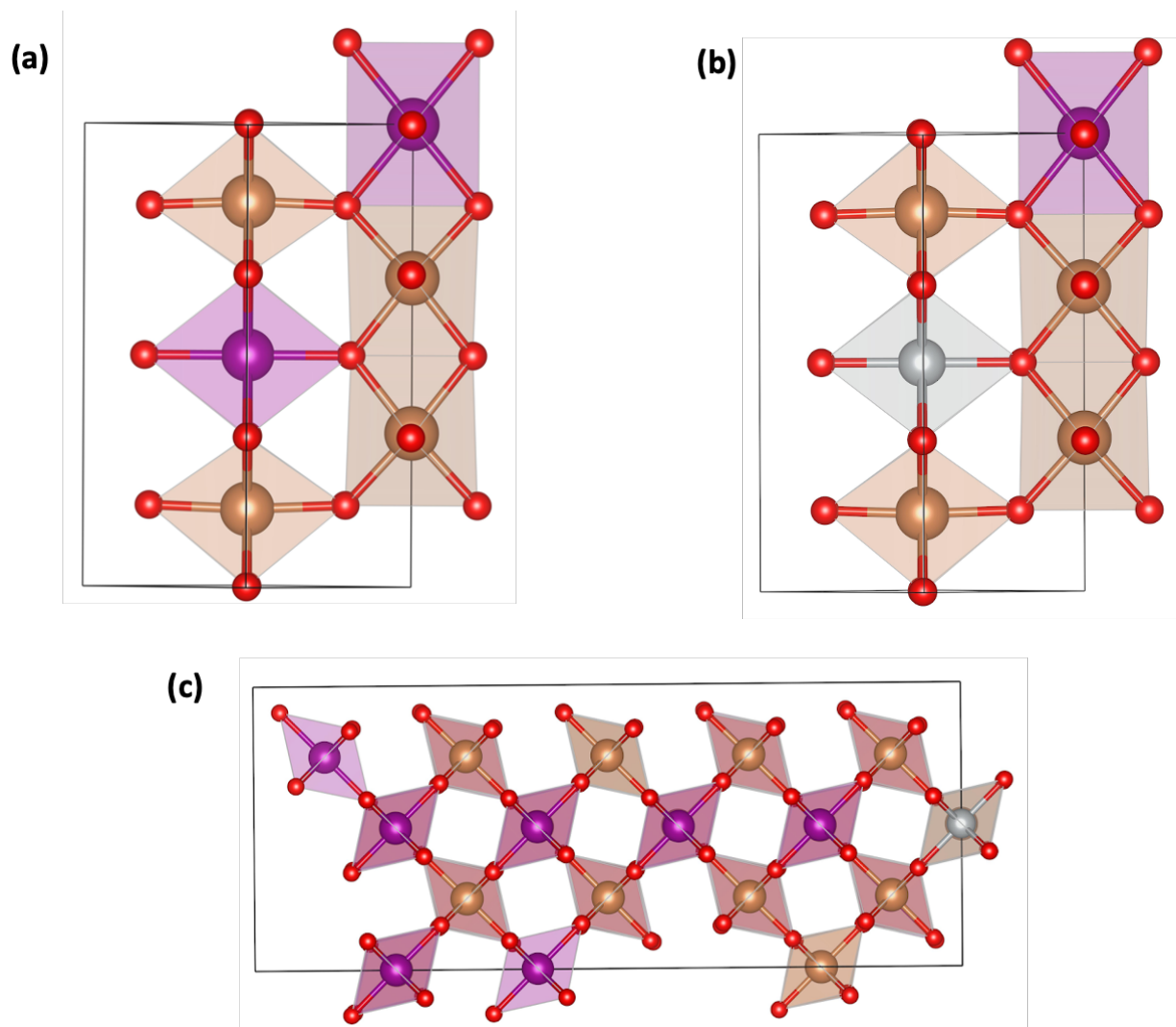
8

Figure 2. DFT optimized bulk structures for a) $(1 \times 1 \times 1)$ $MnSb_2O_6$ and mixed $MnSb_2O_6$ b) doped with 50% transition-metal dopant in a $(1 \times 1 \times 1)$ supercell unit cell and c) 10% transition-metal dopant in a $(2 \times 5 \times 1)$ supercell unit cell. Dopant concentration percentages are shown with respect to the total number of transition-metals in the structure. The primitive bulk unit cell structures are expanded into supercells to achieve the desired dopant concentrations. Mn, O, Sb and the transition-metal dopant are shown in purple, red, brown and gray colors, respectively.

Mixed Antimonates: The experimental dataset of antimonates furthermore consists of mixed $MnSb_2O_6$

with different levels of mixing/doping (e.g., 10, 20, 30, 40, 45 and 50% relative to Mn concentration in the

bulk) with other transition-metals, namely, Ni, Cr and Fe. To simulate these mixed antimonates, Mn in the

host bulk structure or its supercell structure is substituted with stoichiometric amounts of the transition-

metals Ni, Cr and Fe to produce mixed $MnSb_2O_6$ (e.g., 10, 20, 30, 40, 45 and 50% doping) as follows. For

mixed antimonates with doping concentration other than 50%, $(2 \times 5 \times 1)$ supercells of the optimized

MnSb$_2$O$_6$ structure (Figure 1c) are generated. Several positions of these dopant sites and antiferromagnetic ordering are considered, and the lowest energy structure is considered for further electronic and structural analysis. The $(2 \times 5 \times 1)$ supercell consists of 20 Mn sites. As such, 2, 4, 6, 8 and 9 Mn sites are replaced with transition-metal dopants to construct the 10, 20, 30, 40 and 45 % doped antimonates, respectively. For the 50% mixed case, in a $(1 \times 1 \times 1)$ unit cell, which consist of only 2 Mn sites, one of the two Mn sites is replaced with 1 transition-metal dopant. Finally, in consistence with the pure antimonates, an antiferromagnetic (AFM) structure is again considered, and only the lowest energy structure is taken for further analysis. The set of antimonates in the experimental dataset is shown in **Table S1**.

All optimized bulk structures with labels of the corresponding experimental catalyst and the extracted theoretical descriptors from this study are shared at the Catalysis Hub database,[33] and can be accessed directly via https://www.catalysis-hub.org/publications/DeoOxR2023.

2.2.3. Defining Experimental Methodology for Machine Learning

The experimental conditions, measurements and bulk computational electronic and atomic structural descriptors considered for the ML approach are listed in Table 1.

Table 1. Description and shorthand label for experimental and computational descriptors and targets considered for ML modeling. For the computational DFT-based descriptors the distribution over the atoms and/or bonds in the unit cell was considered by including the mean (AVE), maximum (MAX) and the minimum (MIN) as descriptor values.

Shorthand **for ML model**	Defined **experimental terms**
LOAD	**Catalyst loading (mg cm^{-2})** in milligrams of the solid catalyst material present per geometric square centimeter of the electrode deposited onto the working electrode, *i.e.* not including the mass of conductive support.
BNDPCT	**Binder percentage in weight** percentage as calculated by (Nafion mass/(Catalyst mass + Nafion mass)). The Nafion binder is a proton-conducting ionomer that is used to bind together the catalyst layer and carbon support. Nafion is added to the catalyst ink solution alongside catalyst nanoparticles and conductive Vulcan XC-72 carbon support before testing. Nafion is the most commonly used ionomer for industrial fuel cells and necessary for all nanoparticle catalyst systems.
CSUP	**Conductive support to catalyst ratio**: the ratio of mass of the Vulcan XC-72 conductive carbon support to the mass of the catalyst nanoparticles. Conductive supports are commonly used industrially for fuel cells and necessary for almost all nanoparticle catalyst systems.
OPOT	**ORR onset potential**: ML model target, which is taken as a marker of the metric for ORR electrocatalytic activity selected at a cathodic current density of 0.1 mA cm^{-2} as normalized by geometric surface area. The scan rate was kept constant at 20 mV/s.
	Defined **computational terms** (MINimum, MAXimum, and AVErage)
DELEC (DELECMIN, DELECMAX, DELECAVE)	**Number of valence *d* electrons**: The formal number of valence *d* electrons that transition-metal species would have per atom in their bulk metallic phases.
ELNEG (ELNEGMIN, ELNEGMAX, ELNEGAVE)	**Electronegativity (eV)**: Pauling scale of electronegativity for the transition-metals from the antimonates is considered.
ICOHP (ICOHPMIN, ICOHPMAX, ICOHPAVE)	**Integrated Crystal Orbital Hamilton Population (eV)**: The ICOHP is a measure of the bonding interactions between pairs of atoms, including both bonding and antibonding interactions. In this work, the ICOHP is quantified for the metal(*d*) + oxygen(2*p*) interactions for all pairs in the unit cell.
BL (BLMIN, BLMAX, BLAVE)	**Metal-Oxygen Bond length (Ang)**. Bond lengths between all the transition-metal and oxygen pairs in the bulk structures are considered.
O (OMIN, OMAX, OAVE)	**O-2*p* center (eV):** The center of the O-2*p* orbital projected density of states including all oxygen atoms in the unit cell.
D (DMIN, DMAX, DAVE)	**Metal *d*-band center (eV):** The center of the transition-metal *d*-orbital projected density of states, considering all transition-metal atoms in the unit cell.

11

2.3. Data integration and sharing

Our efforts include an integration of the experimental and theoretical data via our database platform Catalysis-hub.org. The experimental data collected from this work are available via the Cathub Python API (https://github.com/SUNCAT-Center/CatHub).

Collected data include experimental characterization spectra (XPS, XRD, XAS), catalyst material properties (composition, morphology, crystal structure and oxidation states), electrochemical sample data (testing conditions and performance) as well as the raw cyclic voltammetry (CV) curves. In addition to the dataset discussed in this work, collected data include oxygen evolution reaction (OER), oxygen reduction reaction (ORR) and hydrogen evolution reaction (HER) catalysis testing data over a broader range of transition-metal compounds. Until now, a public access database for the field of experimental heterogeneous catalysis is lacking and we believe our approach and integrated database will be invaluable for the cataloguing and discovery of novel oxygen reduction and evolution catalysts.

2.4. Machine learning model search

Mathematical equations for ORR onset potentials in the form of genetic programs are evolved using the gplearn code.[46] Modifications to the code used in this project can be found online.[47] These modifications consist of numerical optimization and symbolical simplification of the genetic programs.[28] Furthermore, we enforce that physical quantities of different units can only be added when each multiplied by numerical factors (where the factors then would carry units allowing for addition of the terms). This constraint is important, as the evolved programs are ranked by complexity, and the enforced multiplication with factors adds to the complexity of these programs. During program evolution, the ranking is performed by evaluating the squared error of a model against a random selection of 50% of the transition-metal antimony oxide systems. A penalty term of $\Phi \cdot 10^{-5}$ eV^2 is added to the mean squared error to limit program complexity. Φ is an integer sum measuring program complexity. We consider each mathematical operation (addition, subtraction, multiplication, power, and negation) as well as each integer coefficient and each

variable (descriptor) to increase Φ by one, while floating point numbers as fitting degrees of freedom and divisions are each attributed a weight of two in Φ.

In each generation, the programs compete in k-tournaments,[48] where k programs of the generation are randomly selected (here: $k=10$ out of a population of 200 programs per generation). The program with lowest mean squared error (on a random selection of 50% of the oxides) plus above-described complexity penalty is selected as the fittest for promotion to the next generation after 50% crossover probability with a second k-tournament winner and 3% probability for a mutation of the mathematical terms, and in the remaining 47% of cases, the program moves on to the next generation without modifications. The resulting programs forming the next generation are symbolically simplified and numerically optimized (on a random selection of 50% of the training data) before the next round of k-tournaments. We evolve the model search for 10^5 generations and keep track of the fittest programs for a given complexity for final analysis of the model search as presented in the Results section.

We furthermore analyze the programs post evolution using the Akaike Information Criterion (AIC),[49] where the program fitness is given as:

$$AIC = -2L + 2\Phi. \tag{1}$$

L is the maximum of the log-likelihood of the program and Φ accounts for program complexity as described above. Evaluating Eq. (1) on the programs with lowest error for a given complexity, the AIC provides a measure for the tradeoff between model fit and complexity.

Furthermore, we train Gaussian process regression (GPR) models, using the same inputs and target as used for the genetic programming model search. The Gaussian process is constructed from two squared exponential kernels with different isotropic length scale hyperparameters, as well as a white kernel with a noise level hyperparameter. Kernel hyperparameters are optimized during model training with the L-BFGS-B algorithm[50] in the scikit-learn Python package.[51] Features are standardized by subtracting the mean and scaling to unit variance before model training. To identify the most important features in the GPR model we employed a sequential forward feature selection algorithm implemented in scikit-learn. The best descriptors were chosen based on the lowest root mean square error (RMSE) of the test set, using a 5-fold

13

cross validation across the entire dataset. After identifying the best N-feature model, these features are kept in the set while iterating over all possible N+1-feature models from adding one additional descriptor, until the test error starts to increase with the addition of more features due to overfitting.

3. Results and Discussion

3.1. Machine learning model search

The fittest program of a given complexity from the genetic programming search as well as the best performing GPR models with an increasing number of descriptors as inputs are shown in Figure 3. The fittest programs shown in Figure 3a) are chosen from the evolved generations at each complexity as the models with lowest leave-one-out test error. Only models improving this test error with increasing complexity Φ are considered (no model beyond $\Phi=67$ was found that lowered the test error further). In addition, the corresponding training error is shown. Leave-one-out test and training errors for the GPR model search are shown in Figure 3b). With six descriptors, the test error increased over the best five-descriptor model, and the GPR model search was thus terminated here. Both ML approaches yield models for ORR onsets with test errors below 50 mV with moderate complexity and a small number of descriptors, respectively. Both approaches identify conductive support ratios (CSUP), catalyst loading (LOAD), transition-metal–oxygen bond length (BLMIN,MAX,AVE), and transition-metal electronegativity (ELNEGMIN,MAX,AVE) as important descriptors. The inclusion of experimental descriptors, e.g., CSUP and LOAD, is particularly interesting to note as the impact of loading,[1] ionomers,[52] and conductive supports[53] is commonly studied experimentally, and an extension of models typically relying on descriptors from idealized DFT simulations to these additional descriptors is found here to be of importance for predictions on practical, supported oxide nano-catalysts.

The minimum (and thus optimal) AIC (see Figure S9) was found for the genetic programming model $M_{\Phi=24}$ ≈ 0.4 BLAVE - $2 \cdot 10^{-4}/(0.2$ CSUP + 0.02) + 1.2 LOAD/DELECMAX (full list of models with numerical
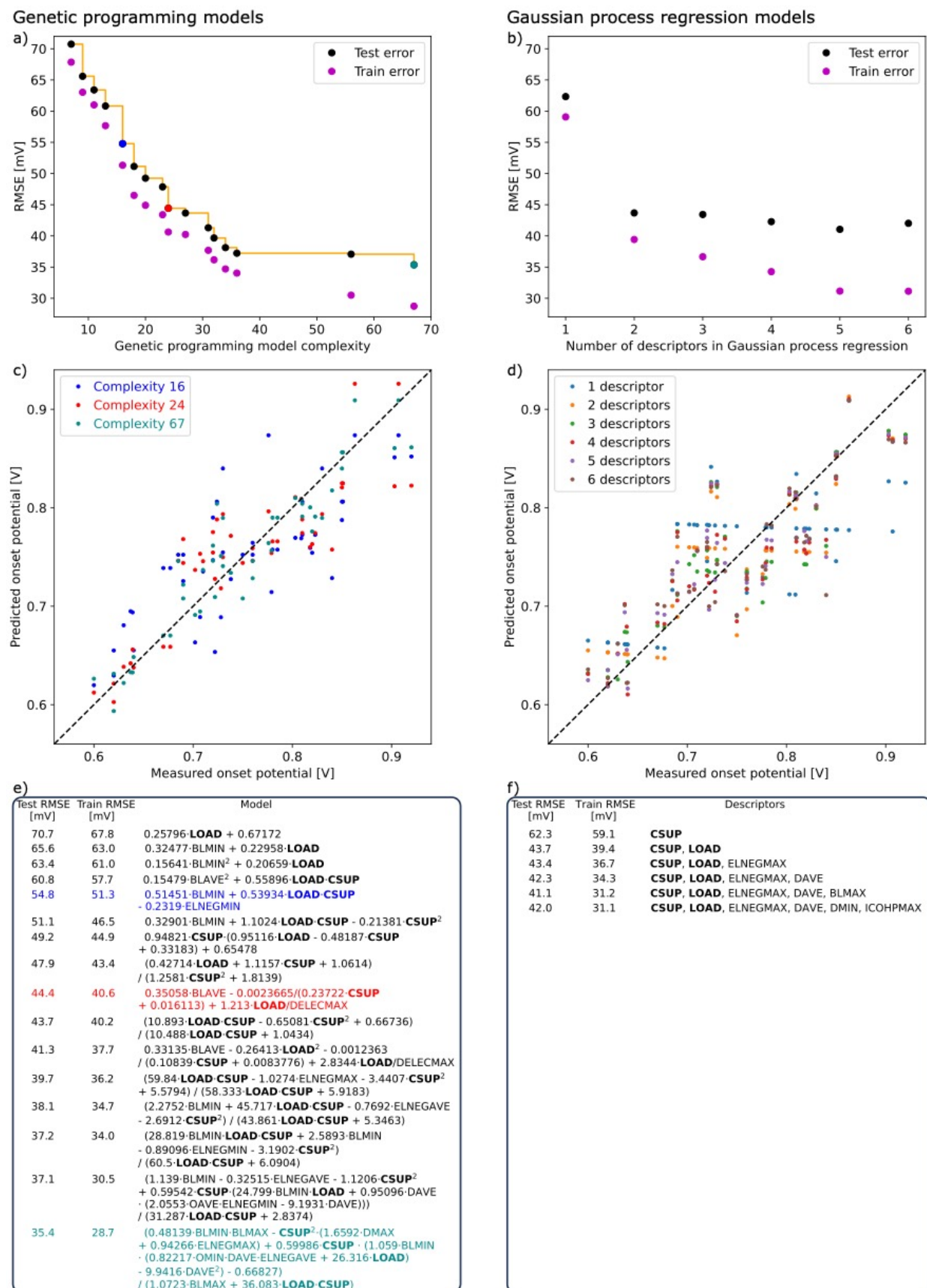
14

Figure 3: ML models for transition-metal antimony oxide ORR onsets found using genetic programming (panels to the left) and using GPR (panels to the right). Shown are the RMSE for genetic programming models (a) and GPR models (b), parity plots for models from the two ML approaches in (c) and (d), and mathematical forms and descriptor inputs are shown in (e) and (f), respectively. Blue, red, and cyan color highlight genetic programming models for which the performance on the experimental data is shown in (c). Bold font face is used for macroscopic experimental catalyst descriptors.

15

coefficients in Figure 3c). Beyond $\Phi=36$, there is a significant increase of the AIC, indicating that these much more complex models are overfits in terms of degrees of freedom. $M_{\Phi=24}$ with optimal AIC, the simple model $M_{\Phi=16} \approx 0.5$ BLMIN + 0.5 LOAD · CSUP - 0.2 ELNEGMIN, for which there is a marked drop in test error from the less complex models, and the most complex model with $\Phi=67$ are highlighted with blue, red, and cyan color, respectively, in a parity plot in Figure 3c), comparing ORR onset predicted from these three models to the measured ones. Similarly, Figure 3d) shows the improved parity from the GPR predictions parameterized with an increasing number of descriptor inputs (employed descriptors shown in Figure 3f).

While both ML approaches lead to similar prediction errors on average, the genetic programming models have the advantage of allowing for direct determination of the influence of the descriptors on the onset potential. The overall predicted increase in onset potential with increasing CSUP can be understood in the following way. With an increased conductive support ratio CSUP, more surface area of the itself semi-conducting nano-particulate antimonate catalyst can be supplied with electrons to run the ORR. While CSUP is not an intrinsic parameter of the antimonate oxide in question, it is an important parameter for the performance of a practical, supported antimonate oxide catalyst.

Overall, an increase in transition-metal–oxygen bond length (BLMIN, BLMAX or BLAVE) is predicted to correlate with an increase in onset potential (decrease in overpotential), while the transition-metal electronegativity (ELNEGMIN, ELNEGMAX, or ELNEGAVE) should be decreased to find such a desired increase of the ORR onset potential. In the following section, we test if newly synthesized antimony oxides doped with lower electronegativity transition-metals would indeed lead to an increase in onset potential over the undoped case.

16

3.4. Experimental testing of model-predicted Sc-doped Ni antimonate oxide ORR performance

Key descriptors provided by ML models can assist with materials design by connecting fundamental material properties to electrocatalyst performance in ways that cannot be easily assessed otherwise. Here, we test the influence of the transition-metal electronegativity on the onset potential. Sc is chosen as a dopant due to both a decrease in electronegativity and an increase in atomic radius compared to the previously studied transition-metal dopants Cr, Mn, Fe, and Ni. Due to the larger atomic radius of Sc, longer transition-metal – oxygen bond lengths are expected, and thus both the electronic and atomic structural descriptors predicted to correlate with the onset potential would be tuned towards an increased onset potential. We choose nickel antimonate as the host material for Sc-doping due to the high electronegativity of Ni and the relatively low onset potential of this oxide, which we aim to improve upon with the Sc-dopant.

A further motivation to choose Sc as a test case here is that Ni-Sc antimonates with $TMSb_2O_6$ stoichiometry are known to exist.[54] Other candidates for dopants with even lower electronegativities and larger atomic radii are elements such as Cs, Ba, La, or Y. These will be interesting to consider. While there are entries in the ICSD[55] for the Ni-Sb-O-Ba and Ni-Sb-O-La systems, these have different stoichiometries[56–63] than the ones considered here and thus do not fit into our focused model search. There are no entries for the cases of Ni-Sb-O-Cs and Ni-Sb-O-Y. Furthermore, the Sc-precursor seemed like a safer, more environmentally friendly option. For example, $Y(NO_3)_3$ is considered category 1 (most severe) for contact and as an aquatic hazard,[64] while $Sc(NO_3)_3$ is only a category 2 oxidizer.[65] For reasons of safety, feasibility of synthesis and stoichiometric compatibility with our model space, we thus excluded these likely promising doping candidates as test cases for our specific models.

Sc doping concentrations of 10% (theoretical, measured value by XPS is 8.2%) and 50% (theoretical, measured value by XPS is 48.4%), substituting for Ni, were considered in the experimental synthesis and characterization. Sc incorporation was achieved by combining scandium nitrate alongside nickel nitrate in the metal salt solution during colloidal synthesis (see Experimental Methods section for more details). The expected Ni: Sc ratios were confirmed by XPS (Figure S10). The high-resolution Ni $2p$ and Sc $2p$ XPS spectra were fit empirically[66,67] to determine the state of each element within the oxide. The Ni $2p$ spectra show similar Ni signatures for the undoped and 10% doped cases, but the Ni is slightly more oxidized in the 50% doped sample. The Sc $2p$ spectra for the two doped samples also indicate similar environments for the 10% and 50% doped materials. XRD is further used to determine the crystal structure of the Ni (Sc) antimonates. The diffraction patterns are consistent with tetragonal $NiSb_2O_6$ (Figure 4a), with and without the addition of Sc doping. The lack of structural change with doping indicates that the Sc intercalation does not disrupt the base oxide lattice and that the material is phase pure.

An ink for each material is prepared by combining 10 mg of solid catalyst material with 2.5 mg of Vulcan XC-72 carbon and 20 microliters of 10% w/w Nafion dispersion in 2 mL pure isopropanol. This results in support-to-catalyst weight ratio of 0.25 and overall binder weight composition of 15%. The mixture is sonicated for 30 minutes prior to use and the ink is deposited via drop-casting to achieve a loading of 0.25 $mg_{catalyst}$ cm^{-2} on a glassy carbon disk of 0.196 cm^2 geometric area. Note that the dispersibility and uniformity of the ink are important factors that are difficult to measure experimentally, but as far as we observed, all three material inks have qualitatively similar dispersion quality. Using an RRDE, all Sc-doped and undoped Ni antimonate oxides were tested under oxygen saturated and nitrogen saturated environment in 0.1 M KOH, ~ pH 13 (Figure 4b). NiSb oxide, which we have previously reported,[4] has an onset of 0.702 ± 0.003 V vs RHE at -0.1 mA cm^{-2}. While Sc incorporation has a detrimental impact on the onset at 10% concentration ($Ni_{0.9}Sc_{0.1}Sb$ oxide, 0.699 ± 0.005 V vs RHE), the higher doping level of 50% leads to a modest improvement of the onset ($Ni_{0.5}Sc_{0.5}Sb$ Oxide, 0.720 ± 0.004 V vs RHE). Notably, Sc doping is seen to slightly suppress selectivity towards H_2O_2.

It is important to note that minimum (MIN) and maximum (MAX) electronic descriptor values stemming from tabulated data (i.e. electronegativities and formal d-electron count per TM ion in bulk) do not depend on the amount of dopant, while the corresponding average values (AVE) do depend on the doping percentage. All DFT-computed MIN,MAX values (band centers and bond-lengths) depend on this degree of doping.

18

Models that only depend on the amount of doping through BLMIN or BLMAX show merely a small increase in onset potential increasing the Sc-doping from 10% to 50% (see Figure 4a). The models with complexities $\Phi=13$ and $\Phi=34$, e.g., depending on BLAVE and ELNEGAVE, respectively, predict a more pronounced increase in onset potential with increasing Sc content (Figure 4a). $M_{\Phi=34}$ predicts an onset potential increase upon 10% Sc-doping by only 4 mV, while 50%-doping is predicted to lead to a 33 mV increase. Given the residual average errors of the models > 30 mV, a mere increase of 4 mV is thus far below the accuracy of the models. Our experimental finding of a slight onset potential decrease for 10% doping is thus well within the uncertainties of our models.

DFT simulations of bulk Sc-doped (10% and 50% doping) nickel antimonate were performed to provide bulk descriptors required as model input. The DFT structural relaxation indeed yielded slightly elongated transition-metal – oxygen bond lengths compared to pure nickel antimonate (by about 2% in the case of 50% doping). Both electronegativity as relevant electronic descriptor as well as these bond lengths as atomic structural descriptor thus should yield an increase in onset potential for the Sc-doped systems.

The performance of the genetic programming and GPR models for the Sc-doped systems is shown in Figure 4c-d. The genetic programming models (Figure 4c) with $\Phi\leq36$ predict onset potentials within at most 0.1V of the experimental results. Models that depend on descriptors that change with dopant concentration (*i.e.* DFT-computed bulk descriptors or average electronegativities) capture the trend of higher Sc dopant concentration leading to higher onset potentials (the remaining models are insensitive to the amount of Sc-doping > 0%). In addition, the significantly more complex models with $\Phi>36$ predict this relative trend correctly, but they predict markedly reduced onset potentials overall for Sc-doping. As is obvious from the AIC for these models (Figure S9), these models likely are an overfit to the training data not containing any Sc-doped systems.

With increasing number of descriptors up to five considered in the GPR (Figure 4d), these models also reach prediction accuracies for the Sc-doped systems below 0.1 V. The models with four and five descriptors as input, which both include the average bulk transition-metal *d*-band center (which depends on the Sc-dopant ratio), moreover, correctly capture the trend of increased onset potential with increased Sc-doping ratio. The model with six descriptors, however, which in the previous section was identified as being close to overfitted due to an increase in test error with respect to the five-descriptor model, incorrectly inverts this trend. The high predicted uncertainty from the more complex models is due to Sc-doped systems being entirely absent from the training data and suggests that such systems should be considered in re-training refined models.
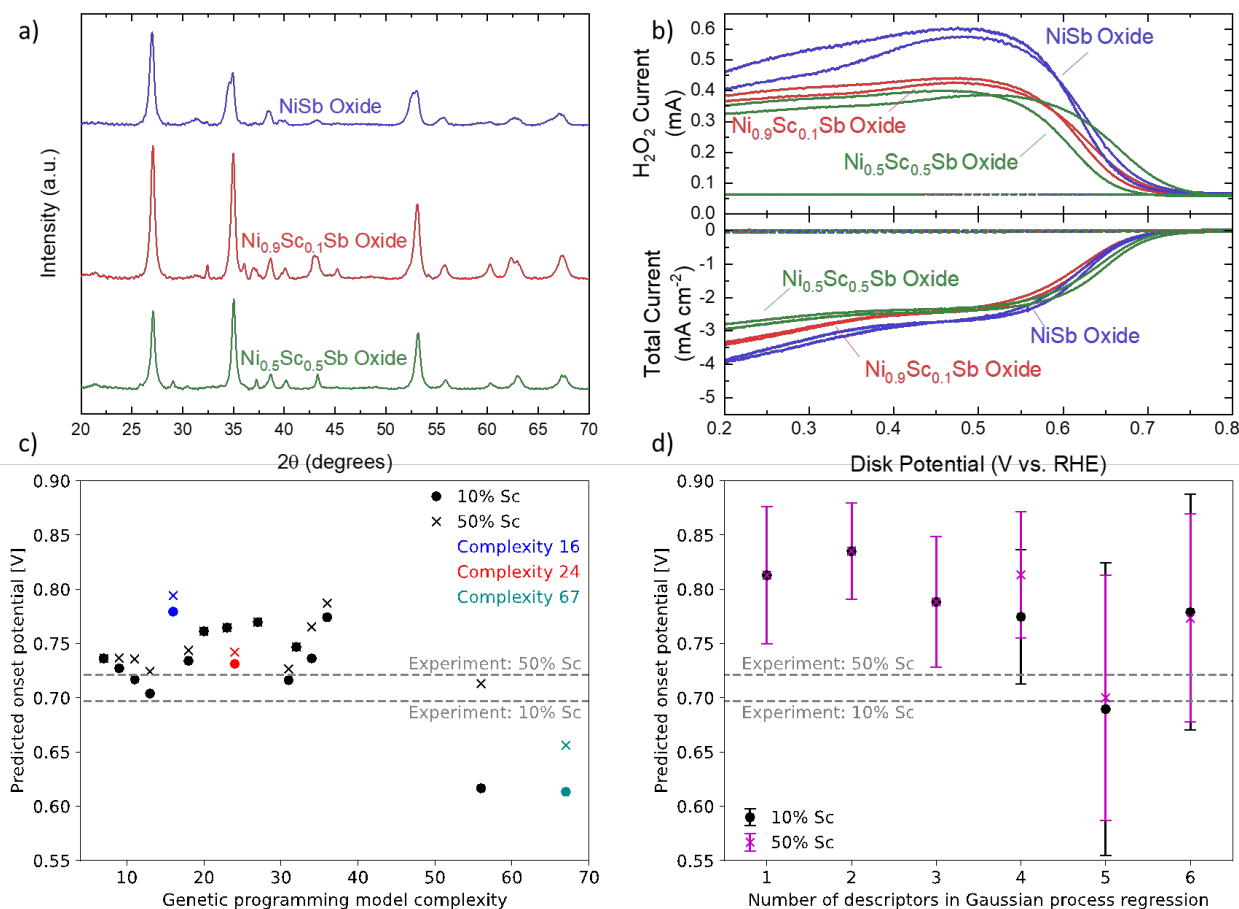
19

Figure 4: (a) Ex situ as prepared XRD structural characterization and (b) RRDE electrochemical evaluation in 0.1 M KOH with O_2 (solid) and N_2 (dotted) of Ni antimonate with 0 (blue), 10 (red), 50 (green) % Sc-doping. Top panel is collection efficiency-normalized ring current with bottom panel displaying total disk current normalized by the geometric surface area. Genetic programming (c) and GPR (d) predictions for ORR onset potentials in alkaline solution for Sc-doped nickel antimonate. The dashed lines show the experimentally observed onset potentials for 10% and 50% Sc doping, respectively. Circles (●) denote the predictions of the ML models for 10% doping and crosses (×) for 50% doping. Blue, red, and cyan color is again used to highlight the genetic programming models considered in more detail in the previous section. Bars denote ± 1 standard deviation predicted from GPR.

4. Conclusions

We performed an ML model search for ORR onset potentials for non-precious transition-metal antimony oxide electrocatalysts. A combination of experimental conditions and measured onset potentials with electronic and atomic structural descriptors computed using DFT for bulk antimony oxide phases served as model input.

20

The ML techniques of genetic programming and Gaussian process regression were successfully employed to derive models predicting ORR onset potentials with about 50 mV test errors. We identified the experimental descriptors, conductive support and loading as having the largest impact on the model. From the human-interpretable mathematical models obtained via genetic programming, we identified the atomic and electronic structural descriptors, transition-metal electronegativities and bond lengths with oxygen as key descriptors to increase the onset potentials via a decrease of the former and an increase of the latter descriptor. We tested this predicted mathematical dependence on these descriptors by synthesizing and testing scandium-doped nickel antimony oxides, where the scandium-dopant introduces the desired descriptor changes. Increased amounts of scandium-doping to were found to increase ORR onsets, in qualitative agreement with predictions for the systems from both the genetic programming and Gaussian process regression approaches.

We believe that integration of simulation data of computationally feasible systems with experimental data on relevant catalyst systems and their analysis through interpretable ML techniques will enable the optimization of catalyst performance through identification of key descriptors and their influence on performance. The importance of experimental parameters above the electronic descriptors in the models for these highly porous carbon-rich systems highlights the general necessity of including such parameters in models for catalyst performance and focusing the comparison or benchmarking of theoretical model descriptor-only models on well defined, low surface area, catalyst systems with consistent loading and mass transport.

5. Data Availability

Our efforts include an integration of these experimental and theoretical data via a web-platform named *cathub* (https://www.catalysis-hub.org/), where both experimental and theoretical catalysis data be publicly accessed.

7. Supporting Information

Data supporting the findings of this study can be found online in the Supporting Information.

8. References

[1]   H. T. Chung, D. A. Cullen, D. Higgins, B. T. Sneed, E. F. Holby, K. L. More, P. Zelenay, *Science* **2017**, *357*, 479–484.

[2]   A. A. Gewirth, J. A. Varnell, A. M. DiAscro, *Chem. Rev.* **2018**, *118*, 2313–2339.

[3]   R. Zeng, Y. Yang, X. Feng, H. Li, L. M. Gibbs, F. J. DiSalvo, H. D. Abruña, *Sci. Adv.* **2022**, *8*, eabj1584.

[4]   G. T. K. K. Gunasooriya, M. E. Kreider, Y. Liu, J. A. Zamora Zeledón, Z. Wang, E. Valle, A.-C. Yang, A. Gallo, R. Sinclair, M. B. Stevens, T. F. Jaramillo, J. K. Nørskov, *ACS Nano* **2022**, *16*, 6334–6348.

[5]   M. E. Kreider, G. A. Kamat, J. A. Zamora Zeledón, L. Wei, D. Sokaras, A. Gallo, M. B. Stevens, T. F. Jaramillo, *J. Am. Chem. Soc.* **2022**, *144*, 22549–22561.

[6]   B. Hammer, J. K. Nørskov, *Surface Science* **1995**, *343*, 211–220.

[7]   Y.-L. Lee, J. Kleis, J. Rossmeisl, Y. Shao-Horn, D. Morgan, *Energy Environ. Sci.* **2011**, *4*, 3966–3970.

[8]   C. F. Dickens, J. H. Montoya, A. R. Kulkarni, M. Bajdich, J. K. Nørskov, *Surface Science* **2019**, *681*, 122–129.

[9]   B. M. Comer, J. Li, F. Abild-Pedersen, M. Bajdich, K. T. Winther, *J. Phys. Chem. C* **2022**, *126*, 7903–7909.

[10] F. Calle-Vallejo, N. G. Inoglu, H.-Y. Su, J. I. Martínez, I. C. Man, M. T. M. Koper, J. R. Kitchin, J. Rossmeisl, *Chem. Sci.* **2013**, *4*, 1245–1249.

[11] M. R. Karim, M. Ferrandon, S. Medina, E. Sture, N. Kariuki, D. J. Myers, E. F. Holby, P. Zelenay, T. Ahmed, *ACS Appl. Energy Mater.* **2020**, *3*, 9083–9088.

[12] N. Madaan, N. R. Shiju, G. Rothenberg, *Catal. Sci. Technol.* **2015**, *6*, 125–133.

[13] Z. M. Shakor, E. N. Al-Shafei, *RSC Adv.* **2023**, *13*, 22579–22592.

[14] P. Uusitalo, A. Sorsa, F. Russo Abegão, M. Ohenoja, M. Ruusunen, *Ind. Eng. Chem. Res.* **2022**, *61*, 4752–4762.

[15] Y. Wang, T.-Y. Chen, D. G. Vlachos, *J. Chem. Inf. Model.* **2021**, *61*, 5312–5319.

[16] J. K. Pedersen, C. M. Clausen, O. A. Krysiak, B. Xiao, T. A. A. Batchelor, T. Löffler, V. A. Mints, L. Banko, M. Arenz, A. Savan, W. Schuhmann, A. Ludwig, J. Rossmeisl, *Angewandte Chemie International Edition* **2021**, *60*, 24144–24152.

[17] K. Okazawa, Y. Tsuji, K. Kurino, M. Yoshida, Y. Amamoto, K. Yoshizawa, *ACS Omega* **2022**, *7*, 45403–45408.

[18] J. A. Esterhuizen, B. R. Goldsmith, S. Linic, *Nat Catal* **2022**, *5*, 175–184.

[19] S.-H. Wang, H. S. Pillai, S. Wang, L. E. K. Achenie, H. Xin, *Nat Commun* **2021**, *12*, 5288.

[20] J. Koza, *Stat. Comput.* **1994**, *4*, 87–112.

[21] D. Angelis, F. Sofos, T. E. Karakasidis, *Arch Computat Methods Eng* **2023**, *30*, 3845–3865.

[22] S.-M. Udrescu, M. Tegmark, *Science Advances* **2020**, *6*, eaay2631.

[23] M. Cranmer, A. Sanchez Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, S. Ho, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., **2020**, pp. 17429–17442.

[24] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L. M. Ghiringhelli, *Phys. Rev. Mater.* **2018**, *2*, 083802.

[25] M. Schmidt, H. Lipson, *Science* **2009**, *324*, 81–85.

[26] Y. Wang, N. Wagner, J. M. Rondinelli, *MRS Communications* **2019**, *9*, 793–805.

[27] K. Takaki, T. Miyao, *Artificial Intelligence in the Life Sciences* **2022**, *2*, 100046.

[28] J. Voss, *J. Phys. Commun.* **2022**, *6*, 035009.

[29] B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, W.-J. Yin, *Nat Commun* **2020**, *11*, 3513.

[30] M. Andersen, K. Reuter, *Acc. Chem. Res.* **2021**, *54*, 2741–2749.

[31] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass, **2006**.

[32] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, *Chem. Rev.* **2021**, *121*, 10073–10141.

[33] K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich, T. Bligaard, *Sci Data* **2019**, *6*, 75.

[34] J. Amador, I. Rasines, *J Appl Crystallogr* **1981**, *14*, 348–349.

[35] H. Guillén-Bonilla, V.-M. Rodríguez-Betancourtt, J. Guillen-Bonilla, L. Gildo-Ortiz, A. Guillen-Bonilla, Y. Casallas-Moreno, O. Blanco-Alonso, J. Reyes-Gómez, *Sensors* **2018**, *18*, 2299.

[36] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169–11186.

[37] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K. W. Jacobsen, *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

23

[38] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[39] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953–17979.

[40] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775.

[41] V. I. Anisimov, F. Aryasetiawan, A. I. Lichtenstein, *J. Phys.: Condens. Matter* **1997**, *9*, 767–808.

[42] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, A. P. Sutton, *Phys. Rev. B* **1998**, *57*, 1505–1509.

[43] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Materials* **2013**, *1*, 011002.

[44] P. E. Blöchl, O. Jepsen, O. K. Andersen, *Phys. Rev. B* **1994**, *49*, 16223–16233.

[45] S. Maintz, V. L. Deringer, A. L. Tchougréeff, R. Dronskowski, *J. Comput. Chem.* **2016**, *37*, 1030–1035.

[46] T. Stephens, "gplearn," can be found under https://gplearn.readthedocs.io/en/stable/, **2023**.

[47] J. Voss, "Fork of gplearn with symbolic and numerical optimization," can be found under https://github.com/vossgroup/gplearn, **2023**.

[48] M. Affenzeller, S. Winkler, S. Wagner, A. Beham, *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications*, CRC Press, Boca Raton, FL, **2009**.

[49] H. Akaike, in *Selected Papers of Hirotugu Akaike* (Eds.: E. Parzen, K. Tanabe, G. Kitagawa), Springer, New York, **1998**, pp. 199–213.

[50] C. Zhu, R. H. Byrd, P. Lu, J. Nocedal, *ACM Trans. Math. Softw.* **1997**, *23*, 550–560.

[51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

[52] R. Jinnouchi, K. Kudo, K. Kodama, N. Kitano, T. Suzuki, S. Minami, K. Shinozaki, N. Hasegawa, A. Shinohara, *Nat Commun* **2021**, *12*, 4956.

[53] M. N. Islam, A. B. Mansoor Basha, V. O. Kollath, A. P. Soleymani, J. Jankovic, K. Karan, *Nat Commun* **2022**, *13*, 6157.

[54] M. Weil, R. Mathieu, P. Nordblad, S. A. Ivanov, *Crystal Research and Technology* **2014**, *49*, 142–151.

[55] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme, *J Appl Crystallogr* **2019**, *52*, 918–925.

[56] P. Köhl, D. Reinen, *Zeitschrift anorg allge Chemie* **1977**, *433*, 81–93.

[57] A. J. Jacobson, A. J. Calvert, *Journal of Inorganic and Nuclear Chemistry* **1978**, *40*, 447–449.

[58] C. Darie, C. Lepoittevin, H. Klein, S. Kodjikian, P. Bordet, C. V. Colin, O. I. Lebedev, C. Deudon, C. Payen, *Journal of Solid State Chemistry* **2016**, *237*, 166–173.

[59] I. Alvarez, *Solid State Ionics* **1996**, *91*, 265–271.

[60] I. Alvarez, M. L. Veiga, C. Pico, *Journal of Alloys and Compounds* **1997**, *255*, 74–78.

[61] D. G. Franco, R. E. Carbonio, G. Nieva, *Journal of Solid State Chemistry* **2013**, *207*, 69–79.

[62] P. D. Battle, S. I. Evers, E. C. Hunter, M. Westwood, *Inorg. Chem.* **2013**, *52*, 6648–6653.

[63] P. D. Battle, M. Avdeev, J. Hadermann, *Journal of Solid State Chemistry* **2014**, *220*, 163–166.

[64] Sigma Aldrich®, "Safety Data Sheet - Yttrium(III) nitrate hexahydrate," can be found under https://www.sigmaaldrich.com/US/en/sds/aldrich/237957, **2024**.

24

[65] Sigma Aldrich®, "Safety Data Sheet - Scandium(III) nitrate hydrate," can be found under https://www.sigmaaldrich.com/US/en/sds/aldrich/325902, **2024**.

[66] M. C. Biesinger, L. W. M. Lau, A. R. Gerson, R. St. C. Smart, *Applied Surface Science* **2010**, *257*, 887–898.

[67] M. C. Biesinger, B. P. Payne, A. P. Grosvenor, L. W. M. Lau, A. R. Gerson, R. St. C. Smart, *Applied Surface Science* **2011**, *257*, 2717–2730.
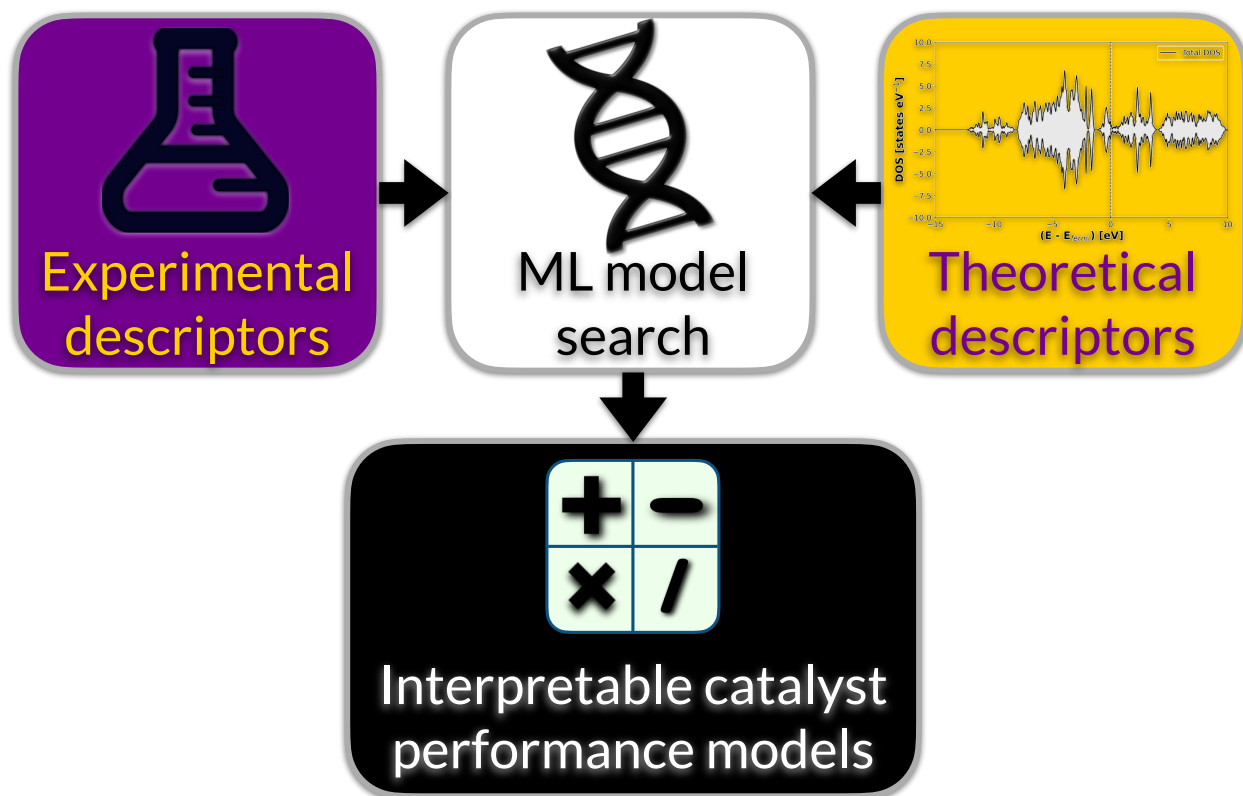
25

TOC Graphic

Table of Contents Graphic: Schematic overview of machine learning approach (ML) combining experimental and theoretical descriptors for discovering human-interpretable models for catalyst performance.